



At the heart of all science is the generation and dissemination of data consistent or inconsistent with specific hypotheses.

Evaluating Scientific Impact

Science rarely moves in an orderly manner. At the heart of all science is the generation and dissemination of data consistent or inconsistent with specific hypotheses, which may be either explicit or implicit. But how do we measure scientific productivity and impact? How do we define and measure success? These questions have been examined by the Committee on Science, Engineering and Public Policy (COSEPUP) of the National Academy of Sciences, the National Academy of Engineering, and the Institute of Medicine in a recent report (1). One of the main findings in this report is that “both applied research and basic research programs supported by the federal government can be evaluated meaningfully on a regular basis.” The report concludes,

Agencies must evaluate their research programs by using measurements that match the character of the research. Differences in the character of the research will lead to differences in the appropriate time scale for measurement, in what is measurable and what is not, and in the expertise needed by those who contribute to the measurement process.

The U.S. Congress and the American public are increasingly interested in the effective management of federal research agencies to ensure that resources provided through the appropriation process are used responsibly and are targeted toward pressing needs. Congress formalized this focus in the Government Performance and Results Act of 1993 (GPRA) (2). The GPRA has three mandates for each government agency: *a*) strategic plans on 5-year cycles, *b*) an annual performance plan, and *c*) an annual performance report. Historically, the National Institutes of Health (NIH) have not been highly scrutinized because the Congress understood that basic research should not be strongly directed and that scientists need freedom to explore unusual phenomena that lead to new discoveries. Both the GPRA and the intense competition for financial support have put the NIH under increasing pressure to fund research with the highest potential impact and to provide money to the most productive scientists.

Research in the environmental health sciences can be classified as basic research, translational or applied research, or population-based research. Basic research focuses on molecular mechanisms of biological processes, for example, by investigating how a particular chemical leads to mutations or how a cell responds to an environmental stress like sunlight. Translational research applies insights derived from basic science to specific models of disease or dysfunction, such as developing a mouse model of neurodegeneration to investigate Parkinson disease. Finally, population-based studies aim to link specific environmental factors with human disease and, if possible, to lessen adverse exposures. The impact in these three areas must include improvement in public health, with population studies yielding beneficial results more rapidly than the other two types of research. However, it is difficult to find the proper metrics and tools to measure impact.

Metrics for evaluating applied or translational research may not be appropriate for basic research (1). For example, research and



William A. Suk, Jerry Phelps, Martha Barnes, and Bennett Van Houten

development divisions in industry can quantify progress in applied research by assessing milestones passed on the way to a specific product. Basic research must be measured in terms of the quality of the science, the relevance to an agency's mission, and the leadership in a field. Basic research often leads to applied, translational, or clinical studies, only many years later. Thus, annual reviews imposed by the GPRA could be detrimental to the progress of basic science unless they are appropriately conceived and applied (1).

Evaluation usually includes five major activities: formulating questions and standards, selecting designs and sampling procedures, collecting information, analyzing information, and reporting information (3). Evaluations should be practical and pragmatic and should reflect the specific needs of an agency, organization, or institution. Evaluations should be qualitative and/or quantitative. Qualitative analysis deals with in-depth descriptions with limited sets of parameters and fewer subjects, whereas quantitative analysis allows much larger numbers of subjects to be analyzed but in limited detail (4). The COSEPUP report on evaluating federal research programs describes a number of methods for analyzing research: bibliometric analysis, economic rate of return, peer review, case study, retrospective analysis, and benchmarking (4). These approaches have particular advantages and disadvantages and run the spectrum from quantitative bibliometric analysis to the more qualitative approach of peer review.

One step in evaluating research is to define the current state of the field. At the heart of scientific evaluation is the quantification of growth and impact. To assess the impact of a particular field, it is important to systematically study the publications in that field. This process, called bibliometric analysis, is powerful and can give quantitative information on specific fields, but it is less useful for comparisons across fields or countries. The use of this tool is complicated because the field of environmental health sciences is so broad; thus, merely counting publications, adding up journal impact factors, and tallying citations is not sufficient. Although these data can be useful, quantifying the impact of one laboratory on a field of study is the true measure of productivity and contribution. Thus, a body of literature that defines progress in the field must be collected and reviewed. This process should help to define the status of the field by crystallizing and articulating current research problems, technical limitations that may be impeding progress, and new research or techniques that might be used to help overcome problems.

Discoveries often flow directly from new tools. The most innovative laboratories develop new approaches or techniques that enable the rest of the field to progress rapidly. Similarly, innovative scientists often make new connections with other fields. However, although trail-blazing is fundamental to the progress of science, scientists who fill in the details about a particular process are also important contributors. Both approaches are essential to science. Thus, evaluations on both qualitative and quantitative levels are needed to answer the following often overlapping questions:

- How does this scientist contribute to the growth of this field?
- Has this laboratory made important contributions that moved the field forward?
- Has the scientist made connections to a new scientific front and uncovered a previously unseen connection?
- Has the scientist developed specific new tools that opened up a new vista of science?
- Does this laboratory publish papers in a timely manner?
- Have these publications had an impact on the field in being cited by others in this field?
- Because the environmental health sciences are multidisciplinary by nature, does this investigator's work contribute to more than one area of study?

One valuable tool in evaluation is a method of tracking all publications emanating from the NIH grants program. Grantees are asked to supply journal reprints with their annual progress report, but they do not always do so. Currently, the best electronic source of NIH-supported publications is MEDLINE, operated by the National Library of Medicine (NLM; Bethesda, MD). If the authors cite grant support, this information is captured in a specific field. Because the current NLM user interface does not allow extensive analysis of publications and grant support, the simple correlation of publications with grant support is difficult. One thorny issue is whether citation indices should also be used to help measure the impact of the specific publications. An entire industry has sprung up around compiling and analyzing such data, for example, the Institute for Scientific Information, which has a large range of products including the Web of Science and the Journal Citation Reports (5), and CHI Research, Inc., which does specific targeted analyses (6). However, citation indices can be misused or even abused (7-9).

An interesting study was conducted in the United Kingdom to evaluate research quality. This study compared the traditional

method of bringing together a working group to carefully review the literature with the use of automated citation indices (10). Conclusions were similar for both the automated literature review and citation indexing and for the more expensive and time-consuming working-group review. If future comparisons validate the use of automated routines, then larger, more complex evaluations can be launched less expensively and more often.

The analytical tools and the outcomes of analyses outlined above will give the NIH increased capability in strategic planning, making it easier for us to comply with the GPRA requirement for a 5-year plan and an annual review. These tools could also enable the institutes to more easily identify areas of research that should be initiated and developed, expanded, or deemphasized as we attempt to address the important environmental public health issues facing the nation.

Bennett Van Houten
Jerry Phelps
Martha Barnes
William A. Suk

Division of Extramural Research and Training
 NIEHS
 Research Triangle Park, North Carolina
 E-mail: vanhout1@niehs.nih.gov

REFERENCES AND NOTES

1. Committee on Science, Engineering, and Public Policy; National Academy of Sciences; National Academy of Engineering; Institute of Medicine. Evaluating Federal Research Programs: Research and the Government Performance and Results Act. Washington, DC:National Academy Press, 1999.
2. Government Performance and Results Act of 1993. Public Law 103-62, 1993.
3. Kosecoff J, Fink A. Evaluation Basics. A Practitioner's Manual. Beverly Hills, CA:Sage Publications, 1982.
4. Patton MQ. How to Use Qualitative Methods in Evaluation. Newbury Park, CA:Sage Publications, 1987.
5. Institute for Scientific Information. Available: <http://www.isinet.com> [cited May 2000].
6. CHI Research, Inc. Available: <http://www.chiresearch.com> [cited May 2000].
7. Pendlebury DA. Science, citation, and funding. *Science* 251:1410-1411 (1991).
8. Seglen PO. Why the impact factor of journals should not be used for evaluating research. *Br Med J* 314:498-502 (1997).
9. Garfield E. Long-term vs. short-term journal impact: does it matter? *Scientist* 12:10-12 (1998).
10. Thomas PR, Watkins DS. Institutional research rankings via bibliometric analysis and direct peer review: a comparative case study with policy implications. *Scientometrics* 41:335-355 (1998).

Upcoming Monographs

2000-2001

- Advances in Uterine Leiomyoma Research
- Mathematical Modeling in Environmental Health Sciences
- Biomedical Research and the Environment
- Reviews in Environmental Health
- Toxicology of Fumonisin
- Nutrition-Toxicology: Evolutionary Aspects

Environmental Health
 PERSPECTIVES
 SUPPLEMENTS